PCR Fusion-Based Approach to Create Reporter Gene Constructs for Expression Analysis in Transgenic *C. elegans*

BioTechniques 32:728-730 (April 2002)

In the nematode *Caenorhabditis elegans*, gene expression patterns are most often determined by generating a fusion of the promoter of the gene under study to a reporter gene such as *lacZ* or *gfp* (3,4,7,8,11). Of the almost 20 000 genes in the worm, the spatial expression pattern of only a few hundred genes has been determined (http://www.wormbase.org). If the reporter gene is fused to the full coding sequence of the gene under study, one can also obtain useful hints about the subcellular localization of the protein. Apart from revealing potential clues about the gene under study, reporter gene fusions serve as invaluable markers to assess the fate of individual cells in defined mutant backgrounds (1,13). The computational comparison of the promoter sequences of large numbers of co-expressed genes will also lead to a better understanding of the underlying logic of transcriptional control (14). Hence, the availability of a large number of reporter genes that provide gene expression information with temporal and spatial resolution is of significant interest in the post-genome era. A method to rapidly create reporter gene fusions on a large scale would thus be a desirable tool to have at hand.

All current methods for generating reporter gene fusions encompass time-consuming DNA subcloning and DNA purification protocols (3,4,8,11). Here a protocol is described to create *gfp* fusion constructs that are ready for injection into the *C. elegans* gonad within one day, with no need for subcloning procedures. The protocol is a modified version of previously described PCR-based fusions of overlapping DNA fragments (9,12) and is schematically outlined in Figure 1. The protocol entails a reaction in which two primary PCR products (Figure 1, product nos. 1 and 2) are fused by PCR with a set of nested primers. In PCR no. 1, the promoter (or the complete gene) under study is amplified from worm genomic DNA that was prepared by digestion with 60 µg/mL proteinase K (1 h lysis at 65ºC; 15 min at 95ºC inactivation; reaction buffer: 10 mM Tris-HCl, pH 8.2, 50 mM KCl, 2.5 mM $MgCl_2$, 0.45% Nonidet™ P-40, 0.45% Tween® 20, 0.01% gelatin) or, alternatively and often more efficiently, from a preparation of cosmid DNA. In the parallel PCR no. 2, the *gfp* coding sequence plus the generically used 3′ untranslated region (UTR) from the *unc-54* gene are amplified from the standard Fire vector pPD95.75 (http://www.ciwemb.edu/pages/firelab.html) (4).

The 3′ primer for the promoter/gene (Figure 1, termed "B") has a 24-nucleotide overhang to the *gfp* vector pPD95.75. If product no. 1 contains the coding region of the gene of interest, then it is important to ensure with primer B that the *gfp* is fused in frame to the gene of interest. Primer "C" does not need to have an overlap to the specific promoter/gene to be amplified, thus making C a generic primer that can be used for every fusion reaction.

Gel purification of the two primary products was often found to inhibit the fusion PCR for unknown reasons. The two primary products are thus used with no further purification for the fusion PCR, which further speeds up the whole process. The concentration of the product # 1 and # 2 is roughly estimated by agarose gel electrophoresis, and an aliquot of the reaction is then diluted with water to 10–50 ng/µL each product. In case the yield of the PCR product is less than 10 ng/µL, it can also be used undiluted; we have even encountered cases in which the first PCR product was invisible on a gel and, nevertheless, obtained a fusion product. After this estimation and dilution step, 1 µL each diluted (or undiluted) reaction is used in the fusion PCR. For this reaction, it is obligatory to use nested primers (Figure 1, A* and D*).

Although in most cases one will get a single band from the PCR fusion, another band can occasionally be seen, possibly a *gfp* dimer; sometimes this additional band may be much stronger than the desired PCR fusion product. This band can be ignored and considered as some sort of "carrier DNA" for the ensuing microinjection into the worms. The concentration of product no. 3 is estimated by agarose gel electrophoresis, and the reaction is diluted

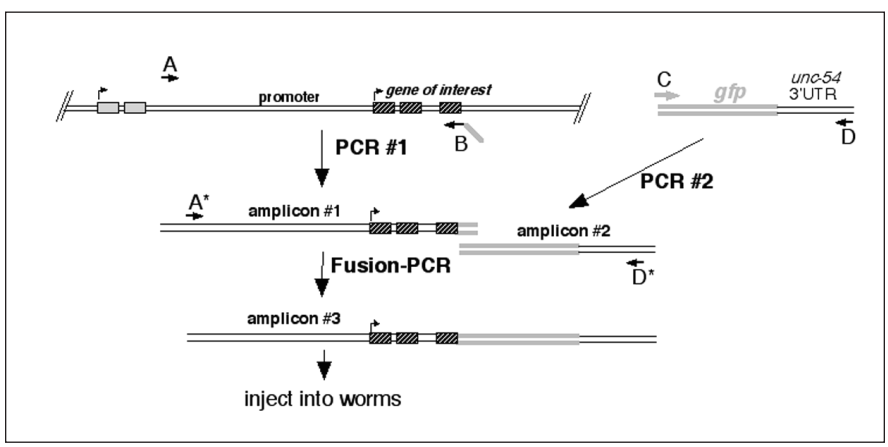

Figure 1. Outline of fusion PCR protocol. The templates are worm genomic DNA (PCR # 1) and the Fire vector pPD95.75 (PCR # 2). Note that the fusion can be made to either the coding sequence of the gene (i.e., translational fusion) or only to the promoter of the gene (i.e., transcriptional fusion). The PCR conditions are: template concentrations, approximately 1 ng/50 µL reaction; primer concentration, 0.3 µM; and nucleotide concentration, 0.2 mM. The primer sequences are: A, 5′ upstream, 20–25 nucleotides long; A*, nested to A (3–10 bp downstream of A); B, spanning 20–24 nucleotides of the end of the gene to fuse + 24 nucleotides of *gfp* vector pPD95.75 (i.e., the sequence of the PLUS strand, 5′-AGCTTGCATGCCTGCAGGTCGACT-3′). For example, if the 3′ end of the gene/promoter has the PLUS strand sequence "5′-AGAGAGAGAGAGAGAGAGAG-3′", the whole primer B would be "5′-AGTCGACCTGCAGGCATGCAAGCTCTCTCTCTCTCTCTCTCT-3′". C, polylinker beginning of pPD95.75, 5′-AGCTTGCATGCCTGCAGGTCGACT-3′; D, at the end of *unc-54* 3′ UTR, 5′-AAGGGCCCGTACGGCCGACTAGTAGG-3′; D*, closely nested to D, 5′-GGAAACAGTTATGTTTGGTATATTGGG-3′.

with water to a final concentration of 20–50 ng/µL. The diluted reaction is then directly injected into the *C. elegans* gonad with no further purification. The DNA is injected either in wild-type N2 animals using *rol-6* as an injection marker at 100 ng/µL concentration (10) or into *pha-1(e1213)ts* using pBX as an injection marker at 100 ng/µL concentration (5). Notably, neither of these injection marker DNAs has any sort of sequence overlap to the co-injected PCR product, yet co-segregation of the injected DNAs has always been found in stable transgenic lines.

For all of these reactions, we used a *Taq/Pwo* DNA polymerase mixture that is provided in the Expand™ Long Template PCR System (Roche Applied Science, Indianapolis, IN, USA). Buffer no. 2 provided in the Expand System (final $MgCl_2$ concentration: 2.25 mM) together with the PCR program that the manufacturer recommends was found to work in most cases. We typically use a 4-kb promoter but have also had success with pieces greater than 10 kb.

Provided that PCR # 1 worked, the fusion PCR never failed. In our experience with more than 50 constructs so far (2,6) (data not shown), we never had problems getting transgenic *C. elegans* lines and, in greater than 95% of the cases, got clearly discernible *gfp* expression. Although rarely found to be a problem, the only significant obstacle in this protocol is getting the long-range PCR # 1 to work. In addition to the obvious candidates of PCR conditions (annealing temperature and reaction buffer), the key parameters to trouble-shoot are the template [either prepare a fresh genomic DNA or request genomic DNA in cosmid clones (available at the Wellcome Trust Sanger Centre, Cambridge, UK)] and the change of PCR primer sequence. Since the location of primer A does not need to be fixed, it can be moved a few dozen base pairs.

As with every PCR application, the final product could contain sequence errors. Since the product is not subcloned but rather consists of a mixture of PCR products that form multicopy arrays in the nucleus (10), errors may only become relevant if they occur very early in the amplification procedure. The presence of the proofreading *Pwo* DNA polymerase should also lead to a minimization of potential sequence errors. Finally, the relatively low information content (i.e., transcription factor binding sites) of 5′ upstream regions suggests a comparatively low potential impact of single-sequence errors.

A rather standard technique, the fusion of two overlapping DNA fragments by PCR (9,12) has been adapted to create reporter gene fusions from genomic worm DNA that can form stable extrachromosomal arrays in transgenic worms. In principle, the scope of this technique can be widened in different ways. PCR fusions could be generated based on three primary products, for example, if one wants to insert *gfp* within a genomic locus (product # 1 = promoter + first part of the coding region + overlap to *gfp*; product # 2, *gfp*; and product # 3, overlap to *gfp* + second part of the coding region + 3′ UTR of the gene of interest). In addition to reporter gene fusions, the technique can be used to express chimeric proteins or to misexpress a gene under the control of a heterologous promoter (i.e., the *gfp* part in product no. 2 is replaced by the gene to be misexpressed), which would allow, because of the ease of the technique, to conduct large-scale misexpression screens for genes that confer a specific phenotype on a given cell.

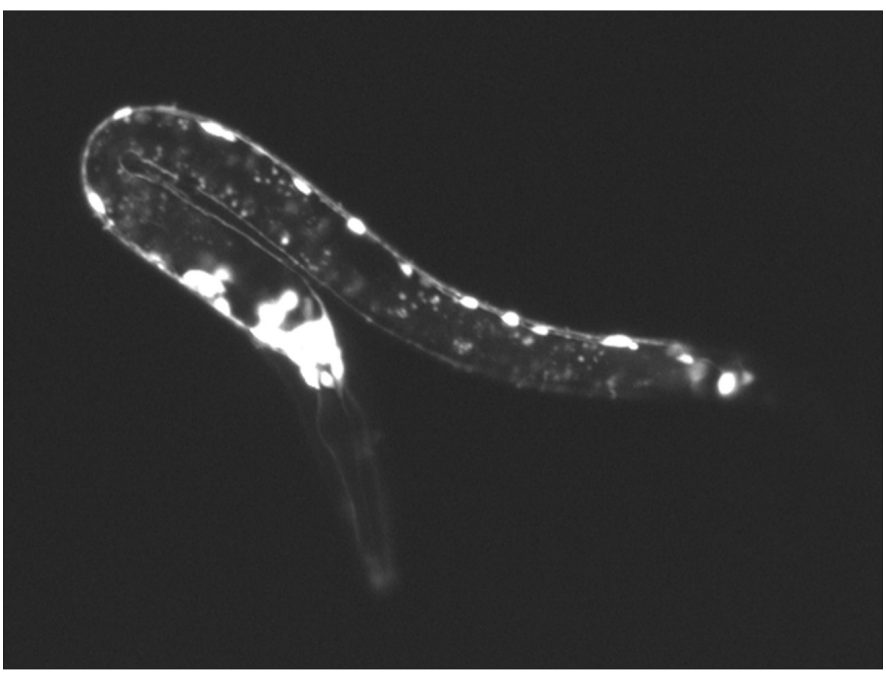

Figure 2. Example of a transgenic *C. elegans* animal expressing a *gfp* reporter gene fusion. The fusion was generated as described in the text and contains 2.5 kb of the promoter of the *C09E7.3* gene. The PCR fragment was injected into *pha-1(e1213)ts* mutant strains. The *gfp* expression can be observed in the nervous system.

REFERENCES

1. **Altun-Gultekin, Z., Y. Andachi, E.L. Tsalik, D. Pilgrim, Y. Kohara, and O. Hobert.** 2001. A regulatory cascade of three homeobox genes, ceh-10, ttx-3, and ceh-23, controls cell fate specification of a defined interneuron class in *C. elegans*. Development *128*:1951-1969.
2. **Aurelio, O., D.H. Hall, and O. Hobert.** 2002. Immunoglobulin-domain proteins required for maintenance of ventral nerve cord organization. Science *295*:686-690.
3. **Chalfie, M., Y. Tu, G. Euskirchen, W.W. Ward, and D.C. Prasher.** 1994. Green fluorescent protein as a marker for gene expression. Science *263*:802-805.
4. **Fire, A., S.W. Harrison, and D. Dixon.** 1990. A modular set of lacZ fusion vectors for studying gene expression in *Caenorhabditis elegans*. Gene *93*:189-198.
5. **Granato, M., H. Schnabel, and R. Schnabel.** 1994. pha-1, a selectable marker for gene transfer in *C. elegans*. Nucleic Acids Res. *22*:1762-1763.
6. **Hobert, O., D.G. Moerman, K.A. Clark, M.C. Beckerle, and G. Ruvkun.** 1999. A conserved LIM protein that affects muscular adherens junction integrity and mechanosensory function in *Caenorhabditis elegans*. J. Cell Biol. *144*:45-57.
7. **Hope, I.A., D.G. Albertson, S.D. Martinelli,**

A.S. Lynch, E. Sonnhammer, and R. Durbin. 1996. The *C. elegans* expression pattern database: a beginning. Trends Genet. *12*:370-371.
8. **Lynch, A.S., D. Briggs, and I.A. Hope.** 1995. Developmental expression pattern screen for genes predicted in the *C. elegans* genome sequencing project. Nat. Genet. *11*:309-313.
9. **McPherson, M.J., P. Quirke, and G.R. Taylor.** 1992. PCR: A Practical Approach, Oxford University Press, Oxford.
10. **Mello, C.C., J.M. Kramer, D. Stinchcomb, and V. Ambros.** 1991. Efficient gene transfer in *C. elegans*: extrachromosomal maintenance and integration of transforming sequences. Embo J. *10*:3959-3970.
11. **Troemel, E.R., J.H. Chou, N.D. Dwyer, H.A. Colbert, and C.I. Bargmann.** 1995. Divergent seven transmembrane receptors are candidate chemosensory receptors in *C. elegans*. Cell *83*:207-218.
12. **White, B.A.** 1993. PCR Protocols. Current Methods and Applications. Humana Press, Totowa, NJ.
13. **Young, J.M. and I.A. Hope.** 1993. Molecular markers of differentiation in *Caenorhabditis elegans* obtained by promoter trapping. Dev. Dyn. *196*:124-132.
14. **Zhang, M.Q.** 1999. Promoter analysis of coregulated genes in the yeast genome. Comput. Chem. *23*:233-250.

I am grateful to Stephen Nurrish for the exchange of ideas, Gary Ruvkun, in whose laboratory I first used this technique, Andy Fire for pPD95.75, and members of my own laboratory, in particular Oscar Aurelio, for repeated use and validation of the technique. Work in my laboratory is funded by a grant from the National Institutes of Health (NS399996-02), the March of Dimes, Whitehall, Searle, Klingenstein, Sloan, Rita Allen, and Irma T. Hirschl Foundations. Address correspondence to Dr. Oliver Hobert, Department of Biochemistry and Molecular Biophysics, Center for Neurobiology and Behavior, Columbia University, College of Physicians and Surgeons, New York, NY 10032, USA. e-mail: or38@columbia.edu

Received 14 November 2001; accepted 18 January 2002.

Oliver Hobert
Columbia University
New York, NY, USA

Partial Heat Denaturation Step during Reverse Transcription and PCR Screening Yields Full-Length 5′-cDNAs

BioTechniques 32:730-736 (April 2002)

RNA secondary and tertiary structures, such as hairpins, stem-loop structures, or even more complicated arrangements like bifurcations and triple-helical elements, play an important role in mRNA stability, RNA processing, and translation efficiency (5). The stability of some of these structural elements often compromises the synthesis of full-length cDNAs (Figure 1). This report describes an intermediate partial heat denaturation step during reverse transcription, followed by the addition of fresh enzyme, which overcomes the problem of reverse transcriptase stoppage. This step is performed at 85°C, which does not allow the perfectly matching RNA-DNA hybrids to denature, whereas potential RNA secondary structures are more likely to melt (Figure 1D). Utilizing this protocol, markedly more full-length cDNAs were recovered compared to other approaches that aim to overcome RNA secondary structures, such as those that use (*i*) the thermostable *r*Th DNA polymerase (Perkin Elmer, Weiterstadt, Germany) that allows reverse transcription of mRNAs at 68°C, or (*ii*) performing reverse transcription at 55°C with non-thermostable reverse transcriptases as recommended by the supplier, and (*iii*) the addition of DMSO [e.g., Moloney murine leukemia virus (MMLV) reverse transcriptase; usage information sheet, Promega, Madison, WI, USA]. The procedure was also successfully applied to primer extension analysis. Furthermore, a rapid PCR screening protocol is described for plasmids containing long inserts, allowing the characterization of full length 5′-cDNAs from total RNA within three days. This method was found to be extremely useful for the detection of low-abundance, full-length cDNA fragments in a pool of shorter fragments, which is a frequently encountered occurrence when RNA integrity is compromised (i.e., when no fresh tissues are available for RNA preparation) or if the target mRNA is transcribed at low levels in the analyzed tissue.

The partial heat denaturation method was successfully applied to clone a variety of cDNAs from different genes. Here, the described studies focused on the isolation of complete 5′-cDNAs of cytochrome *c* oxidase subunits from two species. In several cases in the past, the synthesis of full-length sequences failed because of an unusually high G/C content in the 5′-region, which probably led to unusually stable mRNA secondary structures. Using the thermostable *r*Tth DNA polymerase, MMLV reverse transcriptase (AGS, Heidelberg, Germany) and the SuperScript™ II reverse transcriptase (Invitrogen, Carlsbad, CA, USA) yielded similar unsatisfactory results in the crucial reverse transcription step using standard protocols.

To obtain full-length cDNA sequences, 3′-rapid amplification of cDNA ends (RACE) PCR was first performed as described by Frohman (2), followed by an optimized 5′-RACE protocol, which is described here. In brief, 3′-RACE PCR utilized a dT_{17}-tailed oligonucleotide, Q_T-primer (5′-CCAGTGAGCAGAGTGACGAG-GACTCGAGCTCAAGC$[T]_{17}$-3′), to prime the total cDNA first-strand synthesis, either starting from 5 µg total RNA or 1 µg mRNA. The protocol was identical to the 5′ reverse transcription reaction described below [except that the Q_T-primer is used instead of gene specific primer 1 (GSP1) to prime the cDNA first-strand synthesis]. The appended sequence of the Q_T-primer allowed the utilization of specific primers Q_{outer} (5′-CCAGTGAGCAGAGTGACG-3′) and Q_{inner} (5′-GAGGACTC-GAGCTCAAGC-3′) in subsequent outer and nested PCR amplifications. The 3′-cDNAs were obtained by screening the total cDNA population with a first degenerate primer (e.g., derived from an amino acid sequence or a conserved region from known sequences of other species) using Q_{outer} as counter primer in a standard 50-µL reaction. A nested PCR was performed similarly to increase specificity, using 1